Rombouts, Jeroen V. K.; Bauwens, Luc

Working Paper

Econometrics

Papers, No. 2004,33

Provided in Cooperation with:
CASE - Center for Applied Statistics and Economics, Humboldt University Berlin

Suggested Citation: Rombouts, Jeroen V. K.; Bauwens, Luc (2004) : Econometrics, Papers, No. 2004,33, Humboldt-Universität zu Berlin, Center for Applied Statistics and Economics (CASE), Berlin

This Version is available at:
https://hdl.handle.net/10419/22206

Econometrics

Luc Bauwens[1] and Jeroen V.K. Rombouts[2]

[1] Université catholique de Louvain, Voie du Roman Pays 34, 1348, Louvain-la-Neuve, Belgium bauwens@core.ucl.ac.be
[2] Université catholique de Louvain, Voie du Roman Pays 34, 1348, Louvain-la-Neuve, Belgium rombouts@core.ucl.ac.be

1 Introduction

Since the last decade we live in a digitalized world where many actions in human and economic life are monitored. This produces a continuous stream of new, rich and high quality data in the form of panels, repeated cross-sections and long time series. These data resources are available to many researchers at a low cost. This new era is fascinating for econometricians who can address many open economic questions. To do so, new models are developed that call for elaborate estimation techniques. Fast personal computers play an integral part in making it possible to deal with this increased complexity.

This chapter reviews econometric models for which statistical inference requires intensive numerical computations. A common feature of such models is that they incorporate unobserved (or latent) variables, in addition to observed ones. This often implies that the latent variables have to be integrated from the joint distribution of latent and observed variables. The implied integral is typically of high dimension and not available analytically. Simulation methods are almost always required to solve the computational issue, but they bring new problems. A general introduction on simulation based inference can be found in Gourieroux and Monfort (1997) and Mariano et al. (2000).

The organisation of this chapter is as follows. The first section deals with limited dependent variable models, with a focus on multi-period discrete choice dynamic models. The second section treats the stochastic volatility (SV) model, used in finance and financial econometrics to calibrate the volatility of asset returns, as an alternative to the class of generalized autoregressive conditional heteroskedastic (GARCH) models. It also reviews related dynamic duration models. The last section deals with finite mixture models. Illustrative applications drawn from the recent literature are used. Programs and data are on the web site www.core.ucl.ac.be/econometrics/Bauwens/HBCS/HBCS.htm.

All the models discussed in this chapter are parametric. Nonparametric and semiparametric models may induce additional computational complexity.

We refer to Pagan and Ullah (1999), Horowitz (1998) and Chapter ?? of this volume for examples on these methods.

2 Limited Dependent Variable Models

This section deals with models in which the dependent variable is discrete. Many interesting problems like labour force participation, presidential voting, transport mode choice and brand choice are discrete in nature. In particular, we consider discrete choice models in the case where panel data are available. This allows, for example, to follow individuals with their choices over time, so that richer behavioural models can be constructed. Although the number of parameters in these models does not necessarily increase, the likelihood function, and therefore estimation, becomes more complex. In this section we describe the multinomial multiperiod probit, the multivariate probit and the mixed multinomial logit model. Examples are given.

We refer to Maddala (1983) for a general introduction to limited dependent and qualitative variables in econometrics and to Franses and Paap (2001) for a basic introduction motivating such models in relation to marketing.

2.1 Multinomial Multiperiod Probit

Definition

Denote by U_{ijt} the unobserved utility perceived by individual i who chooses alternative j at time t. This utility may be modelled as follows

$$U_{ijt} = \mathbf{X}_{ijt}^T \boldsymbol{\beta} + \epsilon_{ijt} \tag{1}$$

where $i = 1, \ldots, I,\ j = 1, \ldots, J,\ t = 1, \ldots, T_i$, $\mathbf{X}_{ijt}$ is a k-dimensional vector of explanatory variables, $\boldsymbol{\beta}$ is a k-dimensional parameter vector and ϵ_{ijt} is a random shock known to individual i. This individual chooses alternative j in period t if

$$U_{ijt} > U_{imt} \quad \forall j \neq m. \tag{2}$$

We observe $\mathbf{d}_i = (d_{i1}, \ldots, d_{iT_i})^T$ where $d_{it} = j$ if individual i chooses alternative j at time t. We suppose that there is always only one choice by each individual at each period, i.e. choices are mutually exclusive. The multinomial multiperiod probit model is obtained by assuming

$$\boldsymbol{\epsilon}_i = (\epsilon_{i11}, \ldots, \epsilon_{iJ1}, \ldots, \epsilon_{i1T_i}, \ldots, \epsilon_{iJT_i})^T \sim \text{IIDN}(0, \boldsymbol{\Sigma}) \tag{3}$$

Consequently,

$$P_i = \mathrm{P}(d_i) = \mathrm{P}\left(\bigcap_{m \neq d_{it}} \bigcap_{t=1}^{T_i} U_{i,d_{it},t} > U_{imt}\right)$$
$$= \mathrm{P}\left(\bigcap_{m \neq d_{it}} \bigcap_{t=1}^{T_i} \epsilon_{i,d_{it},t} - \epsilon_{imt} > (\mathbf{X}_{imt} - \mathbf{X}_{i,d_{it},t})^T \boldsymbol{\beta}\right) \quad (4)$$

which is a $(T_i \times J)$-variate integral. However, since individual choices are based on utility comparisons, it is conventional to work in utility differences relative to alternative J. If we multiply the utilities in (1) by a constant, we see that the probability event in (4) is invariant, thus a different scaling of the utilities does not alter the choices of the individuals. The rescaled relative utility is then defined as

$$\begin{aligned}\tilde{U}_{ijt} &= (U_{ijt} - U_{iJt})(\sigma_{11} + \sigma_{JJ} - 2\sigma_{1J})^{-1/2} \\ &= \left((\mathbf{X}_{ijt} - \mathbf{X}_{iJt})^T \boldsymbol{\beta} + \epsilon_{ijt} - \epsilon_{iJt}\right)(\sigma_{11} + \sigma_{JJ} - 2\sigma_{1J})^{-1/2} \\ &= \tilde{\mathbf{X}}_{ijt}^T \boldsymbol{\beta} + \tilde{\epsilon}_{ijt}. \end{aligned} \quad (5)$$

An individual chooses alternative j in period t if

$$\tilde{U}_{ijt} > \tilde{U}_{imt} \quad \forall j \neq m. \quad (6)$$

As an identification restriction, one usually imposes a unit variance for the last alternative expressed in utility differences. Define

$$\tilde{\boldsymbol{\epsilon}}_i = (\tilde{\epsilon}_{i11}, \ldots, \tilde{\epsilon}_{i,J-1,1}, \ldots, \tilde{\epsilon}_{i1T_i}, \ldots, \tilde{\epsilon}_{i,J-1,T_i})^T \sim \mathrm{IIDN}(0, \tilde{\boldsymbol{\Sigma}}) \quad (7)$$

where $\tilde{\boldsymbol{\Sigma}}$ is the transformed $\boldsymbol{\Sigma}$ with $\tilde{\sigma}_{J-1,J-1} = 1$, so that (4) becomes

$$P_i = \mathrm{P}\left(\bigcap_{m \neq d_{it}} \bigcap_{t=1}^{T_i} \tilde{\epsilon}_{i,d_{it},t} - \tilde{\epsilon}_{imt} > (\tilde{X}_{imt} - \tilde{\mathbf{X}}_{i,d_{it},t})^T \boldsymbol{\beta}\right) \quad (8)$$

which is a $T_i(J-1)$-variate integral. Note that when the $\tilde{\epsilon}_{ijt}$'s are serially uncorrelated, this probability event can be calculated by the product of T_i integrals of dimension $J-1$, which is easier to compute but this rules out interesting cases, see the applications below.

Estimation

This section briefly explains how the multinomial multiperiod probit model can be estimated in the classical or Bayesian framework. More details can be found in Geweke et al. (1997).

Classical Estimation

Since we assume independent observations on individuals the likelihood is

$$Pr(\mathbf{d} \mid \mathbf{X}, \boldsymbol{\beta}, \tilde{\boldsymbol{\Sigma}}) = \prod_{i=1}^{I} P_i \tag{9}$$

where $\mathbf{d} = (\mathbf{d}_1, \ldots, \mathbf{d}_I)$ and $\mathbf{X}$ denotes all the observations on the explanatory variables. Evaluation of this likelihood is infeasible for reasonable values of T_i and J. Classical maximum likelihood estimation methods are usually, except in some trivial cases, based on numerical search algorithms that require many times the evaluation of the likelihood function and are therefore not suitable for this model. For more information on classical estimation, see Hajivassiliou and Ruud (1994), Gourieroux and Monfort (1997) and Hajivassiliou and Mc Fadden (1998).

Alternative estimation methods are based on simulations of the choice probabilities. The simulated maximum likelihood (SML) method maximizes the simulated likelihood which is obtained by substituting the simulated choice probabilities in (9). The method of simulated moments is a simulation based substitute for the generalized method of moments. For further information on these estimation methods we refer to Gourieroux and Monfort (1997).

Bayesian Inference

The posterior density is

$$\varphi(\boldsymbol{\beta}, \tilde{\boldsymbol{\Sigma}} \mid \mathbf{d}, \mathbf{X}) \propto Pr(\mathbf{d} \mid \mathbf{X}, \boldsymbol{\beta}, \tilde{\boldsymbol{\Sigma}})\, \varphi(\boldsymbol{\beta}, \tilde{\boldsymbol{\Sigma}}) \tag{10}$$

where $\varphi(\boldsymbol{\beta}, \tilde{\boldsymbol{\Sigma}})$ is the prior density. This does not solve the problem of evaluating a high dimensional integral in the likelihood and it remains hard to compute posterior means for example. Data augmentation, see for example Tanner and Wong (1987), provides a solution because this technique allows to set up a Gibbs sampling scheme using distributions that are easy to draw from. The idea is to augment the parameter vector with $\tilde{\mathbf{U}}$, the latent utilities, so that the posterior density in (10) changes to

$$\varphi(\boldsymbol{\beta}, \tilde{\boldsymbol{\Sigma}}, \tilde{\mathbf{U}} \mid \mathbf{d}, \mathbf{X}) \propto Pr(\mathbf{d} \mid \mathbf{X}, \boldsymbol{\beta}, \tilde{\boldsymbol{\Sigma}}, \tilde{\mathbf{U}})\, f(\tilde{\mathbf{U}} \mid \boldsymbol{\beta}, \tilde{\boldsymbol{\Sigma}})\, \varphi(\boldsymbol{\beta}, \tilde{\boldsymbol{\Sigma}}) \tag{11}$$

implying three blocks in the Gibbs sampler: $\varphi(\boldsymbol{\beta} \mid \tilde{\boldsymbol{\Sigma}}, \tilde{\mathbf{U}}, \mathbf{d}, \mathbf{X})$, $\varphi(\tilde{\boldsymbol{\Sigma}} \mid \boldsymbol{\beta}, \tilde{\mathbf{U}}, \mathbf{d}, \mathbf{X})$ and $\varphi(\tilde{\mathbf{U}} \mid \boldsymbol{\beta}, \tilde{\boldsymbol{\Sigma}}, \mathbf{d}, \mathbf{X})$. For more details on the Gibbs sampler we refer to Chapter **??** and Chapter **??**. For the first two blocks, the model in (5) is the conventional regression model since the utilities, once simulated, are observed. For the last block, remark that $Pr(\mathbf{d} \mid \mathbf{X}, \boldsymbol{\beta}, \tilde{\boldsymbol{\Sigma}}, \tilde{\mathbf{U}})$ is an indicator function since $\tilde{\mathbf{U}}$ is consistent with $\mathbf{d}$ or not.

Applications

It is possible to extend the model in (5) in various ways, such as alternative specific $\boldsymbol{\beta}$'s, individual heterogeneity or a dynamic specification.

Paap and Franses (2000) propose a dynamic specification

$$\Delta\tilde{\mathbf{U}}_{it} = \Delta\tilde{\mathbf{X}}_{it}(\boldsymbol{\alpha}+\boldsymbol{\alpha}_i) + (\boldsymbol{\Pi}-\mathbf{I}_{J-1})\left(\tilde{\mathbf{U}}_{i,t-1} - \tilde{\mathbf{X}}_{i,t-1}(\boldsymbol{\beta}+\boldsymbol{\beta}_i)\right) + \eta_{it} \quad (12)$$

where $\tilde{\mathbf{U}}_{it}$ is the $(J-1)$-dimensional vector of utilities of individual i, $\Delta\tilde{\mathbf{U}}_{it} = \tilde{\mathbf{U}}_{it} - \tilde{\mathbf{U}}_{i,t-1}$, $\tilde{\mathbf{X}}_{i,t-1}$ and $\Delta\tilde{\mathbf{X}}_{it}$ are matrices of dimension $(J-1)\times k$ for the explanatory variables, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are k-dimensional parameter vectors, $\boldsymbol{\Pi}$ is a $(J-1)\times(J-1)$ parameter matrix with eigenvalues inside the unit circle, $\boldsymbol{\eta}_{it} \sim N(0,\tilde{\boldsymbol{\Sigma}})$, and $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ are random individual effects with the same dimension as $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. These individual heterogeneity effects are assumed to be normally distributed: $\boldsymbol{\alpha}_i \sim N(0,\boldsymbol{\Sigma}_{\boldsymbol{\alpha}})$ and $\boldsymbol{\beta}_i \sim N(0,\boldsymbol{\Sigma}_{\boldsymbol{\beta}})$. The specification in (12) is a vector error-correction model where the parameters $\boldsymbol{\alpha}+\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}+\boldsymbol{\beta}_i$ measure respectively the short-run and long-run effects. The parameters in $\boldsymbol{\Pi}$ determine the speed at which deviations from the long-run relationship are adjusted.

The model parameters are $\boldsymbol{\beta}, \boldsymbol{\alpha}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\alpha}_i, \boldsymbol{\beta}_i, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}$ and $\boldsymbol{\Pi}$ and are augmented by the latent utilities $\tilde{U}_{it}$. Bayesian inference may be done by Gibbs sampling as described in the estimation part above. Table 1 describes for each of the nine blocks which posterior distribution is used. For example, β has a conditional (on all other parameters) posterior density that is normal.

Table 1. Summary of conditional posteriors for (12)

parameter	conditional posterior
$\boldsymbol{\beta}$, $\boldsymbol{\beta}_i$, $\boldsymbol{\alpha}$, $\boldsymbol{\alpha}_i$	Multivariate normal distributions
$\tilde{\boldsymbol{\Sigma}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$	Inverted Wishart distributions
$\boldsymbol{\Pi}$	Matrix normal distribution
$\tilde{\mathbf{U}}_{it}$	Truncated multivariate normal

As an illustration we reproduce the results of Paap and Franses (2000), who provided their Gauss code (which we slightly modified). They use optical scanner data on purchases of four brands of saltine crackers. Chintagunta and Honore (1996) use the same data set to estimate a static multinomial probit model. The data set contains all purchases (choices) of crackers of 136 households over a period of two years, yielding 3,292 observations. Variables such as prices of the brands and whether there was a display and/or newspaper feature of the considered brands at the time of purchase are also observed and used as the explanatory variables forming $\mathbf{X}_{ijt}$ (and then transformed into $\tilde{\mathbf{X}}_{ijt}$). Table 2 gives the means of these variables. Display and Feature are

dummy variables, e.g. Sunshine was displayed 13 per cent and was featured 4 per cent of the purchase occasions. The average market shares reflect the observed individual choices, with e.g. 7 per cent of the choices on Sunshine.

Table 2. Means of X_{it} variables in (12)

	Sunshine	Keebler	Nabisco	Private Label
Market share	0.07	0.07	0.54	0.32
Display	0.13	0.11	0.34	0.10
Feature	0.04	0.04	0.09	0.05
Price	0.96	1.13	1.08	0.68

Table 3 shows posterior means and standard deviations for the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ parameters. They are computed from 50,000 draws after dropping 20,000 initial draws. The prior on $\tilde{\boldsymbol{\Sigma}}$ is inverted Wishart, denoted by $IW(\mathbf{S}, \nu)$, with $\nu = 10$ and $\mathbf{S}$ chosen such that $E(\tilde{\boldsymbol{\Sigma}}) = \mathbf{I}_3$. Note that Paap and Franses (2000) use a prior such that $E(\tilde{\boldsymbol{\Sigma}}^{-1}) = \mathbf{I}_3$. For the other parameters we put uninformative priors. As expected, Display and Feature have positive effects on the choice probabilities and price has a negative effect. This holds both in the short run and the long run. With respect to the private label (which serves as reference category), the posterior means of the intercepts are positive except for the first label whose intercept is imprecisely estimated.

Table 3. Posterior moments of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ in (12)

	$\boldsymbol{\beta}$ parameter		*$\boldsymbol{\alpha}$ parameter*			*Intercepts*	
	mean	st. dev.	mean	st. dev.		mean	st. dev.
Display	0.307	(0.136)	0.102	(0.076)	Sunshine	-0.071	(0.253)
Feature	0.353	(0.244)	0.234	(0.090)	Keebler	0.512	(0.212)
Price	-1.711	(0.426)	-2.226	(0.344)	Nabisco	1.579	(0.354)

Table 4 gives the posterior means and standard deviations of $\tilde{\boldsymbol{\Sigma}}$, $\boldsymbol{\Pi}$, $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha}}$. Note that the reported last element of $\tilde{\boldsymbol{\Sigma}}$ is equal to 1 in order to identify the model. This is done, after running the Gibbs sampler with $\tilde{\boldsymbol{\Sigma}}$ unrestricted, by dividing the variance related parameter draws by $\tilde{\boldsymbol{\sigma}}_{J-1,J-1}$. The other parameter draws are divided by the square root of the same quantity. McCulloch et al. (2000) propose an alternative approach where $\tilde{\boldsymbol{\Sigma}}_{J-1,J-1}$ is fixed to 1 by construction, i.e. a fully identified parameter approach. They write

$$\tilde{\boldsymbol{\Sigma}} = \begin{pmatrix} \boldsymbol{\Phi} + \boldsymbol{\gamma}\boldsymbol{\gamma}^T & \boldsymbol{\gamma} \\ \boldsymbol{\gamma}^T & 1 \end{pmatrix} \tag{13}$$

and show that the conditional posterior of $\boldsymbol{\gamma}$ is normal and that of $\boldsymbol{\Phi}$ is Wishart, so that draws of $\tilde{\boldsymbol{\Sigma}}$ are easily obtained. This approach is of particular

Table 4. Posterior means and standard deviations of $\tilde{\mathbf{\Sigma}}$, Π, $\tilde{\mathbf{\Sigma}}_{\boldsymbol{\beta}}$ and $\tilde{\mathbf{\Sigma}}_{\boldsymbol{\alpha}}$ in (12)

$\tilde{\mathbf{\Sigma}}$

$$\begin{pmatrix} 0.563 & -0.102 & 0.433 \\ (0.179) & (0.096) & (0.087) \\ & 0.241 & 0.293 \\ & (0.119) & (0.069) \\ & & 1 \end{pmatrix}$$

$\mathbf{\Pi}$

$$\begin{pmatrix} 0.474 & 0.213 & 0.054 \\ (0.103) & (0.134) & (0.066) \\ 0.440 & 0.685 & -0.196 \\ (0.067) & (0.081) & (0.049) \\ -0.099 & -0.161 & 0.421 \\ (0.091) & (0.138) & (0.087) \end{pmatrix}$$

$\tilde{\mathbf{\Sigma}}_{\boldsymbol{\beta}}$

$$\begin{pmatrix} 0.431 & -0.267 & 0.335 & -0.176 & -0.100 & 0.087 \\ (0.201) & (0.250) & (0.463) & (0.247) & (0.209) & (0.401) \\ & 1.053 & 0.281 & 0.412 & 0.306 & 0.721 \\ & (0.603) & (0.774) & (0.352) & (0.372) & (0.719) \\ & & 5.445 & -1.310 & -1.010 & 0.539 \\ & & (2.268) & (0.999) & (0.853) & (1.120) \\ & & & 1.919 & 1.225 & 1.950 \\ & & & (0.672) & (0.560) & (0.664) \\ & & & & 1.496 & 1.564 \\ & & & & (0.879) & (0.816) \\ & & & & & 4.915 \\ & & & & & (1.319) \end{pmatrix}$$

$\tilde{\mathbf{\Sigma}}_{\boldsymbol{\alpha}}$

$$\begin{pmatrix} 0.207 & -0.023 & -0.004 \\ (0.091) & (0.075) & (0.220) \\ & 0.382 & 0.217 \\ & (0.144) & (0.366) \\ & & 6.672 \\ & & (2.453) \end{pmatrix}$$

interest when a sufficiently informative prior on $\tilde{\mathbf{\Sigma}}$ is used. A drawback of this approach is that the Gibbs sampler has higher autocorrelation and that it is more sensitive to initial conditions.

The relatively large posterior means of the diagonal elements of $\mathbf{\Pi}$ show that there is persistence in brand choice. The matrices $\tilde{\mathbf{\Sigma}}_{\boldsymbol{\beta}}$ and $\tilde{\Sigma}_{\boldsymbol{\alpha}}$ measure the unobserved heterogeneity. There seems to be substantial heterogeneity across the individuals, especially for the price of the products (see the third diagonal elements of both matrices). The last three elements in $\tilde{\mathbf{\Sigma}}_{\boldsymbol{\beta}}$ are related to the intercepts.

The multinomial probit model is frequently used for marketing purposes. For example, Allenby and Rossi (1999) use ketchup purchase data to emphasize the importance of a detailed understanding of the distribution of consumer heterogeneity and identification of preferences at the customer level. In fact, the disaggregate nature of many marketing decisions creates the need for models of consumer heterogeneity which pool data across individuals while allowing for the analysis of individual model parameters. The Bayesian approach is particularly suited for that, contrary to classical approaches that yields only aggregate summaries of heterogeneity.

2.2 Multivariate Probit

The multivariate probit model relaxes the assumption that choices are mutually exclusive, as in the multinomial model discussed before. In that case, d_i may contain several 1's. Chib and Greenberg (1998) discuss classical and Bayesian inference for this model. They also provide examples on voting behavior, on health effects of air pollution and on labour force participation.

2.3 Mixed Multinomial Logit

Definition

The multinomial logit model is defined as in (1), except that the random shock ϵ_{ijt} is extreme value (or Gumbel) distributed. This gives rise to the independence from irrelevant alternatives (IIA) property which essentially means that $Cov(U_{ijt}, U_{ikt}) = 0 \ \forall j, \ \forall k$. Like the probit model, the mixed multinomial logit (MMNL) model alleviates this restrictive IIA property by treating the $\boldsymbol{\beta}$ parameter as a random vector with density $f_{\boldsymbol{\theta}}(\boldsymbol{\beta})$. The latter density is called the mixing density and is usually assumed to be a normal, lognormal, triangular or uniform distribution. To make clear why this model does not suffer from the IIA property, consider the following example. Suppose that there is only explanatory variable and that $\beta \sim \mathrm{N}(\bar{\beta}, \bar{\sigma}^2)$. We can then write (1) as

$$U_{ijt} = X_{ijt}\bar{\beta} + X_{ijt}\bar{\sigma}z + \epsilon_{ijt} \tag{14}$$

$$= X_{ijt}\bar{\beta} + \epsilon^*_{ijt} \tag{15}$$

where $z \sim N(0,1)$, implying that the variance of ϵ^*_{ijt} depends on the explanatory variable and that there is nonzero covariance between utilities for different alternatives.

The mixed logit probability is given by

$$P_i = \int \prod_{t=1}^{T_i} \left(\frac{\mathrm{e}^{\mathbf{X}_{ijt}^T \boldsymbol{\beta}}}{\sum_{j=1}^{J} \mathrm{e}^{\mathbf{X}_{ijt}^T \boldsymbol{\beta}}} \right) f_{\boldsymbol{\theta}}(\boldsymbol{\beta}) \mathrm{d}\boldsymbol{\beta} \tag{16}$$

where the term between brackets is the logistic distribution arising from the difference between two extreme value distributions. The model parameter is $\boldsymbol{\theta}$. Note that one may want to keep elements of $\boldsymbol{\beta}$ fixed as in the usual logit model. One usually keeps random the elements of $\boldsymbol{\beta}$ corresponding to the variables that are believed to create correlation between alternatives. The mixed logit model is quite general. McFadden and Train (2000) demonstrate that any random utility model can be approximated to any degree of accuracy by a mixed logit with appropriate choice of variables and mixing distribution.

Estimation

Classical Estimation

Estimation of the MMNL model can be done by SML or the method of simulated moments or simulated scores. To do this, the logit probability in (16) is replaced by its simulated counterpart

$$SP_i = \frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T_i}\left(\frac{\mathrm{e}^{\mathbf{X}_{ijt}^T\boldsymbol{\beta}^r}}{\sum_{j=1}^{J}\mathrm{e}^{\mathbf{X}_{ijt}^T\boldsymbol{\beta}^r}}\right) \tag{17}$$

where the $\{\boldsymbol{\beta}^r\}_{r=1}^R$ are i.i.d. draws of $f_{\boldsymbol{\theta}}(\boldsymbol{\beta})$. The simulated likelihood is the product of all the individual SP_i's. The simulated log-likelihood can be maximized with respect to $\boldsymbol{\theta}$ using numerical optimization techniques like the Newton-Raphson algorithm. To avoid an erratic behaviour of the simulated objective function for different values of $\boldsymbol{\theta}$, the same sequences of basic random numbers is used to generate the sequence $\{\boldsymbol{\beta}^r\}$ used during all the iterations of the optimizer (this is referred to as the technique of 'common random numbers').

According to Gourieroux and Monfort (1997) the SML estimator is asymptotically equivalent to the ML estimator if T (the total number of observations) and R both tend to infinity and $\sqrt{T}/R \to 0$. In practice, it is sufficient to fix R at a moderate value.

The approximation of an integral like in (16) by the use of pseudo-random numbers may be questioned. Bhat (2001) implements an alternative quasi-random SML method which uses quasi-random numbers . Like pseudo-random sequences, quasi-random sequences, such as Halton sequences, are deterministic, but they are more uniformly distributed in the domain of integration than pseudo-random ones. The numerical experiments indicate that the quasi-random method provides considerably better accuracy with much fewer draws and computational time than does the usual random method.

Bayesian Inference

Let us suppose that the mixing distribution is Gaussian, that is, the vector $\boldsymbol{\beta}$ is normally distributed with mean $\mathbf{b}$ and variance matrix $\mathbf{W}$. The posterior density for I individuals can be written as

$$\varphi(\mathbf{b},\mathbf{W} \mid \mathbf{d},\mathbf{X}) \propto Pr(\mathbf{d} \mid \mathbf{X},\mathbf{b},\mathbf{W})\,\varphi(\mathbf{b},\mathbf{W}) \tag{18}$$

where $Pr(\mathbf{d} \mid \mathbf{X},\mathbf{b},\mathbf{W}) = \prod_{i=1}^{I} P_i$ and $\varphi(\mathbf{b},\mathbf{W})$ is the prior density on $\mathbf{b}$ and $\mathbf{W}$. Sampling from (18) is difficult because P_i is an integral without a closed form as discussed above. We would like to condition on $\boldsymbol{\beta}$ such that the choice probabilities are easy to calculate. For this purpose we augment the model parameter vector with $\boldsymbol{\beta}$. It is convenient to write $\boldsymbol{\beta}_i$ instead of $\boldsymbol{\beta}$ to interpret the random coefficients as representing heterogeneity among

individuals. The $\boldsymbol{\beta}_i$'s are independent and identically distributed with mixing distribution $f(\cdot \mid \mathbf{b}, \mathbf{W})$. The posterior can then be written as

$$\varphi(\mathbf{b}, \mathbf{W}, \boldsymbol{\beta}_I \mid \mathbf{d}, \mathbf{X}) \propto Pr(\mathbf{d} \mid \mathbf{X}, \boldsymbol{\beta}_I)\, f(\boldsymbol{\beta}_I \mid \mathbf{b}, \mathbf{W})\, \varphi(\mathbf{b}, \mathbf{W}) \tag{19}$$

where $\boldsymbol{\beta}_I$ collects the $\boldsymbol{\beta}_i$'s for all the I individuals. Draws from this posterior density can be obtained by using the Gibbs sampler. Table 5 summarizes the three blocks of the sampler.

Table 5. Summary of conditional posteriors for MMNL model

parameter	conditional posterior or sampling method
$\mathbf{b}$	Multivariate normal distribution
$\mathbf{W}$	Inverted Wishart distribution
$\boldsymbol{\beta}_I$	Metropolis Hastings algortihm

For the first two blocks the conditional posterior densities are known and are easy to sample from. The last block is more difficult. To sample from this density, a Metropolis Hastings (MH) algorithm is set up. Note that only one iteration is necessary such that simulation within the Gibbs sampler is avoided. See Train (2003), Chapter 12, for a detailed description of the MH algorithm for the mixed logit model and for guidelines about how to deal with other mixing densities. More general information on the MH algorithm can be found in Chapter **??**.

Bayesian inference in the mixed logit model is called hierarchical Bayes because of the hierarchy of parameters. At the first level, there are the individual parameters $\boldsymbol{\beta}_i$ which are distributed with mean $\boldsymbol{\beta}$ and variance matrix $\mathbf{W}$. The latter are called hyper-parameters, on which we have also prior densities. They form the second level of the hierarchy.

Application

We reproduce the results of McFadden and Train (2000) using their Gauss code available on the web site elsa.berkeley.edu/∼train/software.html. They analyse the demand for alternative vehicles. There are 4,654 respondents who choose among six alternatives (two alternatives run on electricity only). There are 21 explanatory variables among which 4 are considered to have a random effect. The mixing distributions for these random coefficients are independent normal distributions. The model is estimated by SML and uses $R = 250$ replications per observation. Table 6 includes partly the estimation results of the MMNL model. We report the estimates and standard errors of the parameters of the normal mixing distributions, but we do not report the estimates of the fixed effect parameters corresponding to the 17 other explanatory variables.

Table 6. SML estimates of MMNL random effect parameters

Variable	*Mean*		*Standard Deviation*	
Electric vehicle (EV) dummy	-1.032	(0.503)	2.466	(0.720)
Compressed natural gass (CNG) dummy	0.626	(0.167)	1.072	(0.411)
Size	1.435	(0.499)	7.457	(2.043)
Luggage space	1.702	(0.586)	5.998	(1.664)

Robust standard errors within parentheses

For example, the luggage space error component induces greater covariance in the stochastic part of utility for pairs of vehicles with greater luggage space. We refer to McFadden and Train (2000) or Brownstone and Train (1999) for more interpretations of the results.

Train (2003) provides more information and pedagogical examples on the mixed multinomial model.

3 Stochastic Volatility and Duration Models

Stochastic volatility (SV) models may be used as an alternative to generalized autoregressive conditonal heteroskedastic (GARCH) models as a way to model the time-varying volatility of asset returns. Time series of asset returns feature stylized facts, the most important being volatility clustering, which produces a slowly decreasing positive autocorrelation function of the squared returns, starting at a low value (about 0.15). Another stylized fact is excess kurtosis of the distribution (with respect to the Gaussian distribution). See Bollerslev et al. (1994) for a detailed list of the stylized facts and a survey of GARCH models, Shephard (1996) for a comparative survey of GARCH and SV models, and Ghysels et al. (1996) for a survey of SV models focused on their theoretical foundations and their applications in finance. The first four parts of this section deal with SV models while in subsection 3.5 we survey similar models for dynamic duration analysis.

3.1 Canonical SV Model

The simplest version of a SV model is given by

$$\begin{aligned} y_t &= \exp(h_t/2)\, u_t, \qquad u_t \sim N(0,1), \quad t = 1, \ldots, n, \\ h_t &= \omega + \beta h_{t-1} + \sigma v_t, \quad v_t \sim N(0,1), \end{aligned} \tag{20}$$

where y_t is a return measured at t, h_t is the unobserved log-volatility of y_t, $\{u_t\}$ and $\{v_t\}$ are mutually independent sequences, (ω, β, σ) are parameters to be estimated, jointly denoted $\boldsymbol{\theta}$. The parameter space is $\mathbb{R} \times (-1, 1) \times \mathbb{R}_+$. The restriction on β ensures the strict stationarity of y_t. Estimates of β are typically quite close to 1 (in agreement with the first stylized fact), thus β

is a 'persistence' parameter of the volatility. The unconditonal mean of h_t is $\mu = \omega/(1-\beta)$ and the second equation may be parametrized using μ by writing $h_t = \mu + \beta(h_{t-1} - \mu) + \sigma v_t$. Another parametrization removes ω from the second equation while writing the first as $y_t = \tau \exp(h_t/2)\, u_t$ where $\tau = \exp(\omega/2)$. These different parametrizations are in one-to-one correspondance. Which one to choose is mainly a matter of convenience and numerical efficiency of estimation algorithms.

For further use, let y and h denote the $n \times 1$ vectors of observed returns and unobserved log-volatilities, respectively.

3.2 Estimation

Estimation of the parameters of the canonical SV model may be done by the maximum likelihood (ML) method or by Bayesian inference. Other methods have been used but they will not be considered here. We refer to Ghysels et al. (1996), section 5, for a review. ML and, in principle, Bayesian estimation require to compute the likelihood function of an observed sample, which is a difficult task. Indeed, the density of y given $\boldsymbol{\theta}$ and an initial condition h_0 (not explicitly written in the following expressions) requires to compute a multiple integral which has a dimension equal to the sample size:

$$f(y|\boldsymbol{\theta}) = \int f(y, h|\boldsymbol{\theta})\, d\mathbf{h} \tag{21}$$

$$= \int f(y|h, \boldsymbol{\theta}) f(h|\boldsymbol{\theta})\, d\mathbf{h} \tag{22}$$

$$= \int \prod_{t=1}^{n} f(y_t, h_t|\mathbf{Y}_{t-1}, \mathbf{H}_{t-1}, \boldsymbol{\theta})\, d\mathbf{h} \tag{23}$$

where $\mathbf{Y}_t = \{y_i\}_{i=1}^t$ and $\mathbf{H}_t = \{h_i\}_{i=0}^t$. For model (20), this is

$$\int \prod_{t=1}^{n} f_N(y_t|0, e^{h_t}) f_N(h_t|\omega + \beta h_{t-1}, \sigma^2)\, d\mathbf{h}, \tag{24}$$

where $f_N(x|\mu, \sigma^2)$ denotes the normal density function of x, with parameters μ and σ^2. An analytical solution to the integration problem is not available. Even a term by term numerical approximation by a quadrature rule is precluded: the integral of $N(0, \exp(h_n)) \times N(\omega + \beta h_{n-1}, \sigma^2)$ with respect to h_n depends on h_{n-1}, and has to be carried over in the previous product, and so on until h_1. This would result in an explosion of the number of function evaluations. Simulation methods are therefore used.

Two methods directly approximate (23): efficient importance sampling (EIS), and Monte Carlo maximum likelihood (MCML). Another approach, which can only be used for Bayesian inference, works with $f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta})$ as data density, and produces a posterior joint density for $\boldsymbol{\theta}, \mathbf{h}$ given $\mathbf{y}$. The posterior

density is simulated by a Monte Carlo Markov chain (MCMC) algorithm, which produces simulated values of $\boldsymbol{\theta}$ and $\mathbf{h}$. Posterior moments and marginal densities of $\boldsymbol{\theta}$ are then estimated by their simulated sample counterparts. We pursue by describing each method.

EIS (Liesenfeld and Richard (2003))

A look at (24) suggests to sample R sequences $\{h_t^r \sim N(\omega + \beta h_{t-1}, \sigma^2)\}_{t=1}^n$, $r = 1 \dots R$, and to approximate it by $(1/R)\sum_{r=1}^R \prod_{t=1}^n N(0, \exp(h_t^r))$. This direct method proves to be inefficient. Intuitively, the sampled sequences of h_t are not linked to the observations y_t. To improve upon this, the integral (23), which is the convenient expression to present EIS, is expressed as

$$\int \prod_{t=1}^{n} \frac{f(y_t, h_t|\mathbf{Y}_{t-1}, \mathbf{H}_{t-1}, \boldsymbol{\theta})}{m(h_t|\mathbf{H}_{t-1}, \boldsymbol{\phi}_t)} m(h_t|\mathbf{H}_{t-1}, \boldsymbol{\phi}_t)\, d\mathbf{h}, \tag{25}$$

where $\{m(h_t|\mathbf{H}_{t-1}, \boldsymbol{\phi}_t)\}_{t=1}^n$ is a sequence of importance density functions, indexed by parameters $\{\boldsymbol{\phi}_t\}$. These importance functions serve to generate R random draws $\{h_t^1, h_t^2 \dots h_t^R\}_{t=1}^n$, such that the integral is approximated by the sample mean

$$\frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{n} \frac{f(y_t, h_t^r|\mathbf{Y}_{t-1}, \mathbf{H}_{t-1}^r, \boldsymbol{\theta})}{m(h_t^r|\mathbf{H}_{t-1}^r, \boldsymbol{\phi}_t)}. \tag{26}$$

The essential point is to choose the form of $m()$ and its auxiliary parameters $\boldsymbol{\phi}_t$ so as to secure a good match between the product of $m(h_t|\mathbf{H}_{t-1}, \boldsymbol{\phi}_t)$ and the product of $f(y_t, h_t|\mathbf{Y}_{t-1}, \mathbf{H}_{t-1}, \boldsymbol{\theta})$ *viewed as a function of* $\mathbf{h}$. A relevant good match criterion is provided by a choice of $\{\boldsymbol{\phi}_t\}$, for a given family of densities for $m()$, based on the minimization of the Monte Carlo variance of the mean (26). The choice of $\{\boldsymbol{\phi}_t\}$ is explained below, after the choice of $m()$.

A convenient choice for $m()$ is the Gaussian family of distributions. A Gaussian approximation to $f()$, as a function of h_t, given y_t and h_{t-1}, turns out to be efficient. It can be expressed as proportional to $\exp(\phi_{1,t}h_t + \phi_{2,t}h_t^2)$, where $(\phi_{1,t}, \phi_{2,t}) = \boldsymbol{\phi}_t$, the auxiliary parameters. It is convenient to multiply it with $\exp[-0.5\sigma^{-2}(-2m_t h_t + h_t^2 + m_t^2)]$, where $m_t = \omega + \beta h_{t-1}$, which comes from the $N(m_t, \sigma^2)$ term included in $f(y_t, h_t|\mathbf{Y}_{t-1}, \mathbf{H}_{t-1}, \boldsymbol{\theta})$. The product of these two exponential functions can be expressed as a Gaussian density $N(\mu_t, \sigma_t^2)$, where

$$\mu_t = \sigma_t^2(m_t/\sigma^2 + \phi_{1,t}), \quad \sigma_t^2 = \sigma^2/(1 - 2\sigma^2\phi_{2,t}). \tag{27}$$

The choice of the auxiliary parameters can be split into n separate problems, one for each t. It amounts to minimize the sum of squared deviations between $\ln f(y_t|\mathbf{Y}_{t-1}, \mathbf{H}_t^r, \boldsymbol{\theta})$ plus a correction term, see (28), and $\phi_{0,t} + \phi_{1,t}h_t^r + \phi_{2,t}(h_t^r)^2$ where $\phi_{0,t}$ is an auxiliary intercept term. This problem

is easily solved by ordinary least squares. See Liesenfeld and Richard (2003) for a detailed explanation.

Let us summarize the core of the EIS algorithm in three steps (for given $\boldsymbol{\theta}$ and y):
Step 1: Generate R trajectories $\{h_t^r\}$ using the 'natural' samplers $\{N(m_t, \sigma^2)\}$.
Step 2: For each t (starting from $t = n$ and ending at $t = 1$), using the R observations generated in the previous step, estimate by OLS the regression

$$-\frac{1}{2}[h_t^r + y_t^2 e^{-h_t^r} + \left(\frac{\mu_{t+1}^r}{\sigma_{t+1}^r}\right)^2 - \left(\frac{m_{t+1}^r}{\sigma}\right)^2] = \phi_{0,t} + \phi_{1,t} h_t^r + \phi_{2,t}(h_t^r)^2 + \epsilon_t^r \quad (28)$$

where ϵ_t^r is an error term. For $t = n$, the dependent variable does not include the last two terms in the square brackets. The superscript r on μ_{t+1}, σ_{t+1} and m_{t+1} indicates that these quantities are evaluated using the r-th trajectory.
Step 3: Generate R trajectories $\{h_t^r\}$ using the efficient samplers $\{N(\mu_t, \sigma_t^2)\}$ and finally compute (26).

Steps 1 to 3 should be iterated about five times to improve the efficiency of the approximations. This is done by replacing the natural sampler in step 1 by the importance functions built in the previous iteration. It is also possible to start step 1 of the first iteration with a more efficient sampler than the natural one. This is achieved by multiplying the natural sampler by a normal approximation to $f(y_t|h_t, h_{t-1}, \boldsymbol{\theta}) \propto \exp\{-0.5[y_t^2 \exp(-h_t) + h_t]\}$. The normal approximation is based on a second-order Taylor series expansion of the argument of the exponential in the previous expression around $h_t = 0$. In this way, the initial importance sampler links y_t and h_t. This enables one to reduce to three (instead of five) the number of iterations over the three steps. In practical implementations, R can be fixed to 50. When computing (26) for different values of θ, such as in a numerical optimizer, it is important to use common random numbers to generate the set of R trajectories $\{h_t^r\}$ that serve in the computations.

It is also easy to compute by EIS filtered estimates of functions of h_t, such as the conditional standard deviation $\exp(h_t/2)$, conditional on the past returns (but not on the lagged unobserved h_t), given a value of $\boldsymbol{\theta}$ (such as the ML estimate). Diagnostics on the model specification are then obtained as a byproduct: if the model is correctly specified, y_t divided by the filtered estimates of $\exp(h_t/2)$ is a residual that has zero mean, unit variance, and is serially uncorrelated (this also holds for the squared residual).

Richard (1998) contains a general presentation of EIS and its properties.

MCML (Durbin and Koopman (1997))

The likelihood to be computed at $\mathbf{y}$ (the data) and any given $\boldsymbol{\theta}$ is equal to $f(\mathbf{y}|\boldsymbol{\theta})$ and is conveniently expressed as (22) for this method. This quantity is approximated by importance sampling with an importance function defined from an approximating model. The latter is obtained by using the state space representation of the canonical SV model (parametrized with τ):

$$\ln y_t^2 = \ln \tau^2 + h_t + \epsilon_t, \tag{29}$$

$$h_t = \beta h_{t-1} + \sigma v_t. \tag{30}$$

In the canonical SV model, $\epsilon_t = \ln u_t^2$ is distributed as the logarithm of a $\chi^2(1)$ random variable. However the approximating model replaces this with a Gaussian distribution (defined below), keeping the state equation unchanged. Therefore, the whole machinery of the Kalman filter is applicable to the approximating model, which is a Gaussian linear state space model. If we denote by $g(\mathbf{h}|y, \boldsymbol{\theta})$ the importance function that serves to simulate $\mathbf{h}$ (see below), we have

$$f(\mathbf{y}|\boldsymbol{\theta}) = \int \frac{f(\mathbf{y}|\mathbf{h}, \boldsymbol{\theta}) f(\mathbf{h}|\boldsymbol{\theta})}{g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})} \, g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}) \, d\mathbf{h} \tag{31}$$

$$= g(\mathbf{y}|\boldsymbol{\theta}) \int \frac{f(\mathbf{y}|\mathbf{h}, \boldsymbol{\theta})}{g(\mathbf{y}|\mathbf{h}, \boldsymbol{\theta})} \, g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}) \, d\mathbf{h}, \tag{32}$$

where the second equality results from $g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}) g(\mathbf{y}|\boldsymbol{\theta}) = g(\mathbf{y}|\mathbf{h}, \boldsymbol{\theta}) g(\mathbf{h}|\boldsymbol{\theta})$ and $g(\mathbf{h}|\boldsymbol{\theta}) = f(\mathbf{h}|\boldsymbol{\theta})$. All the densities $g(.)$ and $g(.|.)$ are defined from the approximating Gaussian model. In particular, $g(\mathbf{y}|\boldsymbol{\theta})$ is the likelihood function of the Gaussian linear state space model and is easy to compute by the Kalman filter (see the appendix to Sandman and Koopman (1998) for all details). Likewise, $g(\mathbf{y}|\mathbf{h}, \boldsymbol{\theta})$ obtains from the Gaussian densities $g(\ln y_t^2|h_t, \theta)$ resulting from (29) with $\epsilon_t \sim N(a_t, s_t^2)$ where a_t and s_t^2 are chosen so that $g(\mathbf{y}|\mathbf{h}, \boldsymbol{\theta})$ is as close as possible to $f(\mathbf{y}|\mathbf{h}, \boldsymbol{\theta})$. The parameters a_t and s_t^2 are chosen so that $\ln g(\ln y_t^2|\hat{h}_t, \boldsymbol{\theta})$ and $\ln f(\ln y_t^2|\hat{h}_t, \boldsymbol{\theta})$ have equal first and second derivatives, where $\hat{h}_t$ is the smoothed value of h_t provided by the Kalman filter applied to the approximating model. Remark that this is a different criterion from that used in EIS. Finally, $g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$ can be simulated with the Gaussian simulation smoother of de Jong and Shephard (1995).

In brief, the likelihood function is approximated by

$$g(y|\boldsymbol{\theta}) \frac{1}{R} \sum_{r=1}^{R} \frac{f(\mathbf{y}|\mathbf{h}^r, \boldsymbol{\theta})}{g(\mathbf{y}|\mathbf{h}^r, \boldsymbol{\theta})} \tag{33}$$

where $\mathbf{h}^r = \{h_t^r\}_{t=1}^T$ is simulated independently R times with the importance sampler and $g(\mathbf{y}|\boldsymbol{\theta})$ is computed by the Kalman filter. Equation (32) and (33) shows that importance sampling serves to evaluate the departure of the actual likelihood from the likelihood of the approximating model. R is fixed to 250 in practice.

For SML estimation, the approximation in (33) is transformed in logarithm. This induces a bias since the expectation of the log of the sample mean is not the log of the corresponding integral in (32). The bias is corrected by adding $s_w^2/(2R\bar{w})$ to the log of (33), where s_w^2 is the sample variance of the ratios $w^r = f(\mathbf{y}|\mathbf{h}^r, \boldsymbol{\theta})/g(\mathbf{y}|\mathbf{h}^r, \boldsymbol{\theta})$ and $\bar{w}$ is the sample mean of the same ratios, i.e. $\bar{w}$ is the sample mean appearing in (33). Moreover, Durbin and Koopman

(1997) use antithetic and control variables to improve the efficiency of the estimator of the log-likelihood function.

Durbin and Koopman (2000) present several generalizations of MCML (e.g. the case where the state variable in non-Gaussian) and develop analogous methods for Bayesian inference.

MCMC (Kim et al. (1998))

We present briefly the 'Mixture Sampler', one of the three algorithms added by Kim et al. (1998) to the six algorithms already in the literature at that time (see their paper for references). They approximate the density of $\epsilon_t = \ln u_t^2$ by a finite mixture of seven Gaussian densities, such that in particular the first four moments of both densities are equal. The approximating density can be written as

$$f_a(\epsilon_t) = \sum_{i=1}^{7} \Pr[s_t = i] f(\epsilon_t | s_t = i) = \sum_{i=1}^{7} \Pr[s_t = i] f_N(\epsilon_t | b_i - 1.2704, c_i^2) \quad (34)$$

where s_t is a discrete random variable, while $\Pr[s_t = i]$, b_i and c_i are known constants (independent of t). The constant -1.2704 is the expected value of a $\ln \chi^2(1)$ variable.

The crux of the algorithm is to add $\mathbf{s} = \{s_t\}_{t=1}^n$ to $\boldsymbol{\theta}$ and $\mathbf{h}$ in the MCMC sampling space. This makes it possible to sample $\mathbf{h}|\mathbf{s}, \boldsymbol{\theta}, \mathbf{y}$, $\mathbf{s}|\mathbf{h}, \mathbf{y}$ and $\boldsymbol{\theta}|\mathbf{h}, \mathbf{y}$ within a Gibbs sampling algorithm. Remark that $\mathbf{s}$ and $\boldsymbol{\theta}$ are independent given $\mathbf{h}$ and $\mathbf{y}$. Moreover, $\mathbf{h}$ can be sampled entirely as a vector. The intuition behind this property is that, once $\mathbf{s}$ is known, the relevant term of the mixture (34) is known for each observation, and since this is a Gaussian density, the whole apparatus of the Kalman filter can be used. Actually, this a bit more involved since the relevant Gaussian density depends on t, but an augmented Kalman filter is available for this case.

Sampling $\mathbf{h}$ as one block is a big progress over previous algorithms, such as in Jacquier et al. (1994), where each element h_t is sampled individually given the other elements of $\mathbf{h}$ (plus $\boldsymbol{\theta}$ and $\mathbf{y}$). The slow convergence of such algorithms is due to the high correlations between the elements of $\mathbf{h}$.

Kim et al. (1998) write the model in state space form, using μ rather than ω or τ as a parameter, i.e.

$$\ln y_t^2 = h_t + \epsilon_t, \quad (35)$$

$$h_t - \mu = \beta(h_{t-1} - \mu) + \sigma v_t. \quad (36)$$

The 'Mixture Sampler' algorithm is summarized in Table 7. Notice that once $\boldsymbol{\theta}$ has been sampled, it is easy to transform the draws of μ into equivalent draws of ω or τ by using the relationships between these parameters. Since inference is Bayesian, prior densities must be specified. For σ^2, an inverted gamma prior density is convenient since the conditional posterior is in the same class and

Table 7. Summary of 'Mixture Sampler' algorithm

parameter	conditional posterior or sampling method
h	Gaussian simulation smoother
s	Univariate discrete distribution for each s_t
σ^2	Inverted gamma distribution
β	Rejection or Metropolis-Hastings sampler
μ	Normal distribution

easy to simulate. For β, any prior can be used since the conditional posterior is approximated and rejection sampling is used. A beta prior density is advocated by Kim et al. (1998). For μ, a Gaussian or uninformative prior results in a Gaussian conditional posterior.

Kim et al. (1998) also propose an algorithm to compute filtered estimates of h_t, from which model diagnostics can be obtained as described above for EIS.

3.3 Application

For illustration, estimates of the canonical SV model parameters are reported in Table 8 for a series of 6,107 centred daily returns of the Standard and Poor's 500 (SP500) composite price index (period: 02/01/80-30/05/03, source: Datastream). Returns are computed as 100 times the log of the price ratios. The sample mean and standard deviation are equal to 0.03618 and 1.0603, respectively.

Table 8. ML and Bayesian estimates of SV model (20)

	EIS (ω)	MCML (τ)	MCMC (τ)
ω/τ	-0.00524 (0.00227)	0.863 (0.0469)	0.864 (0.0494)
β	0.982 (0.00385)	0.982 (0.00389)	0.983 (0.00382)
σ	0.149 (0.0138)	0.147 (0.0135)	0.143 (0.0139)
llf	-8023.98	-8023.80	
time	2.36 min.	7.56 min.	6.23 min.
code	Gauss	Ox	Ox

llf: value of log-likelihood function at the reported estimate; EIS, MCML, and MCMC are defined in subsection 3.2

We used computer codes provided by the authors cited above. For EIS, we received the code from R. Liesenfeld, for MCML and MCMC we downloaded them from the web site staff.feweb.vu.nl/koopman/sv.

For SML estimation by EIS or MCML, identical initial values ($\beta = 0.96$, $\sigma = 0.15, \omega = 0.02$ or $\tau = 0.01$) and optimization algorithms (BFGS) are used,

but in different programming environments. Therefore, the computing times are not fully comparable, although a careful rule of thumb is that Ox is two to three times faster than Gauss (see Cribari-Neto (1997)). Reported execution times imply that EIS appears to be at least six times faster than MCML. This is a reversal of a result reported by Sandman and Koopman (1998) (p 289), but they compared MCML with a precursor of EIS implemented by Danielson (1994). More importantly, the two methods deliver quasi-identical results.

MCMC results are based on 18,000 draws after dropping 2,000 initial draws. The posterior means and standard deviations are also quite close to the ML results. The posterior density of σ (computed by kernel estimation) is shown in Figure 1 together with the large sample normal approximation to the density of the ML estimator using the EIS results. The execution time for MCMC is difficult to compare with the other methods since it depends on the number of Monte Carlo draws. It is however quite competitive since reliable results are obtained in no more time than MCML in this example.

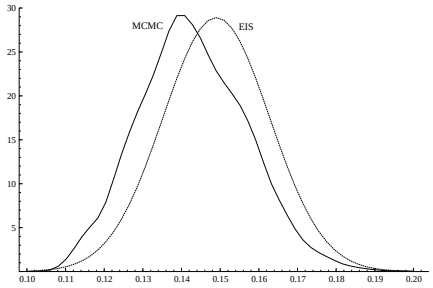

Fig. 1. Posterior density of σ and normal density of the MLE

3.4 Extensions of the Canonical SV Model

The canonical model presented in (20) is too restrictive to fit the excess kurtosis of many return series. Typically, the residuals of the model reveal that the distribution of u_t has fatter tails than the Gaussian distribution. The assumption of normality is most often replaced by the assumption that $u_t \sim t(0, 1, \nu)$, which denotes Student-t distribution with zero mean, unit variance, and degrees of freedom parameter $\nu > 2$. SML estimates of ν are usually between 5 and 15 for stock and foreign currency returns using daily data. Posterior means are larger because the posterior density of has a long tail to the right.

Several other extensions of the simple SV model presented in (20) exist in the literature. The mean of y_t need not be zero and may be a function of explanatory variables x_t (often a lag of y_t and an intercept term). Similarly h_t may be a function of observable variables (z_t) in addition to its own lags. An extended model along these lines is

$$\begin{aligned} y_t &= x_t^T \gamma + \exp(h_t/2)\, u_t, \\ h_t &= \omega + z_t^T \alpha + \beta h_{t-1} + \sigma v_t, \end{aligned} \tag{37}$$

It should be obvious that all these extensions are very easy to incorporate in EIS (see Liesenfeld and Richard (2003)) and MCML (see Sandman and Koopman (1998)). Bayesian estimation by MCMC remains quite usable but becomes more demanding in research time to tailor the algorithm for achieving a good level of efficiency of the Markov chain (see Chib et al. (2002), in particular p 301-302, for such comments).

Chib et al. (2002) also include a jump component term $k_t q_t$ in the conditional mean part to allow for irregular, transient movements in returns. The random variable q_t is equal to 1 with unknown probablity κ and zero with probability $1-\kappa$, whereas k_t is the size of the jump when it occurs. These time-varying jump sizes are assumed independent draws of $\ln(1+k_t) \sim N(-0.5\delta^2, \delta^2)$, δ being an unknown parameter representing the standard deviation of the jump size. For daily SP500 returns (period: 03/07/1962-26/08/1997) and a Student-t density for u_t, Chib et al. (2002) report posterior means of 0.002 for κ, and 0.034 for δ (for prior means of 0.02 and 0.05, respectively). This implies that a jump occurs on average every 500 days, and that the variability of the jump size is on average 3.4 per cent. They also find that the removal of the jump component from the model reduces the posterior mean of ν from 15 to 12, which corresponds to the fact that the jumps capture some outliers.

Another extension consists of relaxing the restriction of zero correlation (ρ) between u_t and v_t. This may be useful for stock returns for which a negative correlation corresponds to the leverage effect of the financial literature. If the correlation is negative, a drop of u_t, interpreted as a negative shock on the return, tends to increase v_t and therefore h_t. Hence volatility increases more after a negative shock than after a positive shock of the same absolute value, which is a well-known stylized fact. Sandman and Koopman (1998) estimate such a model by MCML, and report $\hat{\rho} = -0.38$ for daily returns of the SP500 index (period: 02/01/80-30/12/87), while Jacquier et al. (2004) do it by Bayesian inference using MCMC and report a posterior mean of ρ equal to -0.20 on the same data. They use the same reparametrization as in (13) to impose that the first diagonal element of the covariance matrix of u_t and σv_t must be equal to 1. This covariance matrix is given by

$$\Sigma = \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix} = \begin{pmatrix} 1 & \psi \\ \psi & \phi^2 + \psi^2 \end{pmatrix} \tag{38}$$

where the last matrix is a reparametrization. This enables to use a normal prior on the covariance ψ and an inverted gamma prior on ϕ^2, the conditional variance of σv_t given u_t. The corresponding conditional posteriors are of the same type, so that simulating these parameters in the MCMC algorithm is easy. This approach can also be used if u_t has a Student-t distribution.

Multivariate SV models are also on the agenda of researchers. Liesenfeld and Richard (2003) estimate by EIS a one-factor model introduced by Shephard (1996), using return series of four currencies. Kim et al. (1998), section 6.6, explain how to deal with the multi-factor model case by extending the MCMC algortihm reviewed in subsection 3.2.

3.5 Stochastic Duration and Intensity Models

Models akin to the SV model have been used for dynamic duration analysis by Bauwens and Veredas (2004) and Bauwens and Hautsch (2003). The context of application is the analysis of a sequence of time spells between events (also called durations) occurring on stock trading systems like the New York Stock Exchange (NYSE). Time stamps of trades are recorded for each stock on the market during trading hours every day, resulting in an ordered series of durations. Marks, such as the price, the exchanged quantity, the prevailing bid and ask prices, and other observed features may also be available, enabling to relate the durations to the marks in a statistical model. See Bauwens and Giot (2001) for a presentation of the issues.

Let $0 = t_0 < t_1 < t_2 < \ldots < t_n$ denote the arrival times and $d_1, d_2 \ldots d_n$ denote the corresponding durations, i.e. $d_i = t_i - t_{i-1}$. The stochastic conditional duration (SCD) model of Bauwens and Veredas (2004) is defined as

$$\begin{aligned} d_i &= \exp(\psi_i)\, u_i, & u_i &\sim D(\gamma), \quad t = 1, \ldots, n, \\ \psi_i &= \omega + \beta\psi_{i-1} + \sigma v_i, & v_i &\sim N(0,1), \end{aligned} \tag{39}$$

where $D(\gamma)$ denotes some distribution on the positive real line, possibly depending on a parameter γ. For example, Bauwens and Veredas use the Weibull distribution and the gamma distribution (both with shape parapeter denoted by γ).Assuming that the distribution of u_i is parameterized so that $E(u_i) = 1$, ψ_i is the logarithm of the unobserved mean of d_i, and is modelled by a Gaussian autoregressive process of order one. It is also assumed that $\{u_i\}$ and $\{v_i\}$ are mutually independent sequences. The parameters to be estimated are $(\omega, \beta, \sigma, \gamma)$, jointly denoted $\boldsymbol{\theta}$. The parameter space is $\mathbb{R} \times (-1, 1) \times \mathbb{R}_+ \times \mathbb{R}_+$.

The similarity with the canonical SV model (20) is striking. A major difference is the non-normality of u_i since this is by definition a positive random variable. This feature makes it possible to identify γ. Therefore, the estimation methods available for the SV model can be adapted to the estimation of SCD models. Bauwens and Veredas (2004) have estimated the SCD model by the quasi-maximum likelihood (QML) method, since the first equation of the model may be expressed as $\ln d_i = \psi_i + \ln u_i$. If $\ln u_i$ were Gaussian, the model would be a Gaussian linear state space model and the Kalman filter could be directly applied. QML relies on maximizing the likelihood function as if $\ln u_i$ were Gaussian. The QML estimator is known to be consistent but inefficient relative to the ML estimator which would obtain if the correct distribution of $\ln u_i$ were used to define the likelihood function. Galli (2003), Chapter 3,

has studied by simulation the loss of efficiency of QML relative to ML. ML estimation assuming a Weibull distribution is done by applying the EIS algorithm. For a sample size of 500 observations, the efficiency loss ranges from 20 to 50 per cent, except for the parameter ω, for which it is very small. He also applied the EIS method using the same data as in Bauwens and Veredas (2004). For example, for a dataset of 1,576 volume durations of the Boeing stock (period: September-November 1996; source: TAQ database of NYSE), the ML estimates are: $\hat{\omega} = -0.028$, $\hat{\beta} = 0.94$, $\hat{\sigma}^2 = 0.0159$, $\hat{\gamma} = 1.73$. They imply a high persistence in the conditional mean process (corresponding to duration clustering), a Weibull distribution with an increasing concave hazard function, and substantial heterogeneity. Notice that an interesting feature of the SCD model is that the distribution of u_i conditional to the past information, but marginalized with respect to the latent process, is a Weibull mixed by a lognormal distribution.

Strickland et al. (2003) have designed a MCMC algortithm for the SCD model (39) assuming a standard exponential distribution for u_i. The design of their MCMC algorithm borrows features from Koopman and Durbin's MCML approach and one of the MCMC algorithms used for the SV model.

As an alternative to modelling the sequence of durations, Bauwens and Hautsch (2003) model directly the arrival times through the intensity function of the point process. Their model specifies a dynamic intensity function, where the intensity function is the product of two intensity components: an observable component that depends on past arrival times, and a latent component. The logarithm of the latter is a Gaussian autoregressive process similar to the second equation in (20) and (39). The observable component may be a Hawkes process (Hawkes (1971)) or an autoregressive intensity model (Russell (1999)). When the model is multivariate, there is an observable intensity component specific to each particular point process, while the latent component is common to all particular processes. Interactions between the processes occur through the observable components and through the common component. The latter induces similar dynamics in the particular processes, reflecting the impact of a common factor influencing all processes. Bauwens and Hautsch use intensity-based likelihood inference, with the EIS algorithm to deal with the latent component.

4 Finite Mixture Models

Many econometric isuues require models that are richer or more flexible than the conventional regression type models. Several possibilities exist. For example, as explained in subsection 2.3, the logit model is made more realistic by generalizing it to a mixed logit. Many models currently used in econometrics can be generalized in such a way.

In this section, we assume that the univariate or multivariate observations $\mathbf{y}_j$ are considered as draws of

$$\tilde{f}(y_j) = \sum_{g=1}^{G} \eta_g f(\mathbf{y}_j|\boldsymbol{\theta}_g) \tag{40}$$

with $\eta_1 + \ldots + \eta_G = 1$. The densities $f(\cdot|\boldsymbol{\theta}_g)$ are called component distributions. The observation $\mathbf{y}_j$ comes from one of these component distributions but we do not observe to which component it belongs. The mixture problem involves making inference about the η_g's and the parameters of the component distributions given only a sample from the mixture. The closer the component distributions are to each other, the more difficult this is because of problems of identifiability and computational instability.

4.1 Inference and Identification

The structure of (40) implies that the likelihood for all the J observations contains G^J terms

$$L(\boldsymbol{\eta}, \boldsymbol{\theta}|y) \propto \prod_{j=1}^{J} \left(\sum_{g=1}^{G} \eta_g f(\mathbf{y}_j|\boldsymbol{\theta}_g) \right) \tag{41}$$

where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_G)^T$ and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_G)^T$ contain all the parameters and $\mathbf{y}$ denotes all the data. Maximum likelihood estimation using numerical optimization techniques, requiring many evaluations of the likelihood function, becomes cumbersome, if not unfeasible, for large G and J. This is even worse for multivariate observations.

Bayesian inference on finite mixture distributions by MCMC sampling is explained in Diebolt and Robert (1994). Gibbs sampling on $(\boldsymbol{\eta}, \boldsymbol{\theta})$ is difficult since the posterior distributions of $\boldsymbol{\eta}|\boldsymbol{\theta}, \mathbf{y}$ and $\boldsymbol{\theta}|\boldsymbol{\eta}, \mathbf{y}$ are generally unknown. For the same reason as for the probit model in section 2.1 and the stochastic volatility model in section 3, inference on the finite mixture model is straightforward once the state or group of an observation is known. Data augmentation is therefore an appropriate way to render inference easier. Define the state indicator S_j which takes value $s_j = g$ when $\mathbf{y}_j$ belongs to state or group g where $g \in \{1, \ldots, G\}$. Denote by $\mathbf{S}$ the J-dimensional discrete vector containing all the state indicators. To facilitate the inference, prior independence, that is $\varphi(\boldsymbol{\eta}, \boldsymbol{\theta}, \mathbf{S}) = \varphi(\boldsymbol{\eta})\varphi(\boldsymbol{\theta})\varphi(\mathbf{S})$, is usually imposed. As shown in the next examples, the posterior distributions $\mathbf{S}|\boldsymbol{\eta}, \boldsymbol{\theta}, \mathbf{y}$, $\boldsymbol{\theta}|\boldsymbol{\eta}, \mathbf{S}, \mathbf{y}$ and $\boldsymbol{\eta}|\boldsymbol{\theta}, \mathbf{S}, \mathbf{y}$ are either known distributions easy to sample from or they are distributions for which a second, but simpler, MCMC sampler is set up. A Gibbs sampler with three main blocks may therefore be used.

The complete data likelihood of the finite mixture is invariant to a relabeling of the states. This means that we can take the labeling $\{1, 2, \ldots, G\}$ and do a permutation $\{\rho(1), \rho(2), \ldots, \rho(G)\}$ without changing the value of the likelihood function. If the prior is also invariant to relabeling then the posterior has this property also. As a result, the posterior has potentially $G!$ different

modes. To solve this identification or label switching problem, identification restrictions have to be imposed.

Note that the inference described here is conditional on G, the number of components. There are two modelling approaches to take care of G. First, one can treat G as an extra parameter in the model as is done in Richardson and Green (1997) who make use of the reversible jump MCMC methods. In this way, the prior information on the number of components can be taken explicitly into account by specifying for example a Poisson distribution on G in such a way that it favors a small number of components. A second approach is to treat the choice of G as a problem of model selection. By so-doing one separates the issue of the choice of G from estimation with G fixed. For example, one can take $G = 2$ and $G = 3$ and do the estimation separately for the two models. Then Bayesian model comparison techniques (see Chapter **??**) can be applied, for instance by the calculation of the Bayes factor, see Cowles and Carlin (1996) and Chib (1995) for more details.

4.2 Examples

We review two examples. The first example fits US quarterly GNP data using a Markov switching autoregressive model. The second example is about the clustering of many GARCH models.

Markov Switching Autoregressive Model

Frühwirth-Schnatter (2001) uses US quarterly real GNP growth data from 1951:2 to 1984:4. This series was initially used by Hamilton (1989) and is displayed in Figure 2. The argument is that contracting and expanding periods

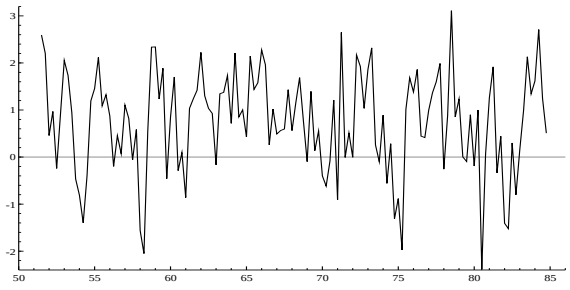

Fig. 2. US real GNP growth data in percentages (1951:2 to 1984:4)

are generated by the same model but with different parameters. These models are called state- (or regime-) switching models.

After some investigation using Bayesian model selection techniques, the adequate specification for the US growth data is found to be the two-state switching AR(2) model

$$y_t = \beta_{i,1} y_{t-1} + \beta_{i,2} y_{t-2} + \beta_{i,3} + \epsilon_{t,i} \qquad \epsilon_{t,i} \sim \mathrm{N}(0, \sigma_i^2) \qquad i = 1, 2. \quad (42)$$

The idea behind the choice of two states is motivated by the contracting (negative growth) and expanding periods (positive growth) in an economy. The conditional posteriors for the σ_i^2's are independent inverted gamma distributions. For the β_i's, the conditional posteriors are independent normal distributions. Inference for the switching model in (42) is done in two steps. The first step is to construct an MCMC sample by running the random permutation sampler. Generally speaking, a draw from the random permutation sampler is obtained as follows:

(i) Draw from the model by use of the Gibbs sampler for example.
(ii) Relabel the states randomly.

By so-doing, one samples from the unconstrained parameter space with balanced label switching. Note that in (ii), there are $G!$ possibilities to relabel when there are G possible states.

In the second step, this sample is used to identify the model. This is done by visual inspection of the posterior marginal and bivariate densities. Identification restrictions need to be imposed to avoid multimodality of the posterior densities. Once suitable restrictions are found, a final MCMC sample is constructed to obtain the moments of the constrained posterior density. The latter sample is constructed by permutation sampling under the restrictions, which means that (ii) is replaced by one permutation defining the constrained parameter space.

In the GNP growth data example, two identification restrictions seem possible, namely $\beta_{1,1} < \beta_{2,1}$ and $\beta_{1,3} < \beta_{2,3}$, see Frühwirth-Schnatter (2001) for details. Table 9 provides the posterior means and standard deviations of the $\beta_{i,j}$'s for both identification restrictions.

Table 9. Posterior means and standard deviations

	$\beta_{1,1} < \beta_{2,1}$		$\beta_{1,3} < \beta_{2,3}$	
	Contraction	Expansion	Contraction	Expansion
$\beta_{i,1}$	0.166 (0.125)	0.33 (0.101)	0.249 (0.164)	0.295 (0.116)
$\beta_{i,2}$	0.469 (0.156)	-0.129 (0.093)	0.462 (0.164)	-0.114 (0.098)
$\beta_{i,3}$	-0.479 (0.299)	1.07 (0.163)	-0.557 (0.322)	1.06 (0.175)

The GNP growth in contraction and expansion periods not only have different unconditional means, they are also driven by different dynamics. Both identification restrictions result in similar posterior moments.

Clustering of Many GARCH Models

Bauwens and Rombouts (2003) focus on the differentiation between the component distributions via different conditional heteroskedasticity structures by the use of GARCH models. In this framework, the observation y_j is multivariate and the $\boldsymbol{\theta}_g$'s are the parameters of GARCH(1,1) models. The purpose is to estimate many, of the order of several hundreds, GARCH models. Each financial time series belongs to one of the G groups but it is not known a priori which series belongs to which cluster.

An additional identification problem arises due to the possibility of empty groups. If a group is empty then the posterior of $\boldsymbol{\theta}_g$ is equal to the prior of $\boldsymbol{\theta}_g$. Therefore an improper prior is not allowed for $\boldsymbol{\theta}_g$. The identification problems are solved by using an informative prior on each $\boldsymbol{\theta}_g$. The identification restrictions use the fact that we work with GARCH models: we select rather non-overlapping supports for the parameters, such that the prior $\varphi(\boldsymbol{\theta}) = \prod_{g=1}^{G} \varphi(\boldsymbol{\theta}_g)$ depends on a labeling choice. Uniform prior densities on each parameter, on finite intervals, possibly subject to stationarity restrictions, are relatively easy to specify.

Bayesian inference is done by use of the Gibbs sampler and data augmentation. Table 10 summarizes the three blocks of the sampler.

Table 10. Summary of conditional posteriors

parameter	conditional posterior or sampling method
$\mathbf{S}$	Multinomial distribution
$\boldsymbol{\eta}$	Dirichlet distribution
$\boldsymbol{\theta}$	Griddy-Gibbs sampler

Because of the prior independence of the $\boldsymbol{\theta}_g$'s, the griddy-Gibbs sampler is applied separately G times.

As an illustration we show the posterior marginals of the following model

$$\tilde{f}(y_j) = \sum_{g=1}^{3} \eta_g f(y_j|\boldsymbol{\theta}_g) \tag{43}$$

with $\eta_1 = 0.25$, $\eta_2 = 0.5$, $J = 100$ and $T_j = 1000$. The components are defined more precisely as

$$f(y_j|\boldsymbol{\theta}_g) = \prod_{t=1}^{T_j} f(y_{j,t}|\boldsymbol{\theta}_g, I_{j,t}) \tag{44}$$

$$y_{j,t}|\boldsymbol{\theta}_g, I_{j,t} \sim N(0, h_{j,t}) \tag{45}$$

$$h_{j,t} = (1 - \alpha_g - \beta_g)\tilde{\omega}_j + \alpha_g (y_{j,t-1})^2 + \beta_g h_{j,t-1} \tag{46}$$

where $I_{j,t}$ is the information set until $t-1$ containing (at least) $y_{j,1}, \ldots, y_{j,t-1}$ and initial conditions which are assumed to be known. For the simulation of the data $\tilde{\omega}_j$ is fixed equal to one which implies that the unconditional variance for every generated data series is equal to one. However, the constant $\tilde{\omega}_j$ in the conditional variance is not subject to inference, rather it is fixed at the empirical variance of the data. Table 11 presents the true values, the prior intervals on the $\boldsymbol{\theta}_g$'s and posterior results on $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$.

Table 11. Posterior results on $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$ ($G = 3$)

		η_1	η_2	η_3
True value		**0.25**	**0.50**	**0.25**
Mean		0.2166	0.4981	0.2853
Standard deviation		0.0555	0.0763	0.0692
Correlation matrix		1	-0.4851	-0.2677
		-0.4851	1	-0.7127
		-0.2677	-0.7127	1
		$g = 1$	$g = 2$	$g = 3$
True value	α_g	**0.04**	**0.12**	**0.20**
	β_g	**0.90**	**0.60**	**0.40**
Prior interval	α_g	0.001,0.07	0.07,0.15	0.15,0.25
	β_g	0.65,0.97	0.45,0.75	0.20,0.60
Mean	α_g	0.0435	0.1041	0.1975
	β_g	0.8758	0.5917	0.4369
Standard deviation	α_g	0.0060	0.0092	0.0132
	β_g	0.0238	0.0306	0.0350
Correlation α_g, β_g		-0.7849	-0.71409	-0.7184

Bauwens and Rombouts (2003) succesfully apply this model to return series of 131 US stocks. Comparing the marginal likelihoods for different models, they find that $G = 3$ is the appropriate choice for the number of component distributions.

Other interesting examples of finite mixture modelling exist in the literature. Frühwirth-Schnatter and Kaufmann (2002) develop a regime switching panel data model. Their purpose is to cluster many short time series to capture asymmetric effects of monetary policy on bank lending. Deb and Trivedi (1997) develop a finite mixture negative binomial count model to estimate six measures of medical care demand by the elderly. Chib and Hamilton (2000) offer a flexible Bayesian analysis of the problem of causal inference in models with non-randomly assigned treatments. Their approach is illustrated using hospice data and hip fracture data.

References

Allenby, G. and Rossi, P. (1999). Marketing models of consumer heterogeneity. *Journal of Econometrics*, 89:57–78.

Bauwens, L. and Giot, P. (2001). *Econometric Modelling of Stock Market Intraday Activity*. Kluwer.

Bauwens, L. and Hautsch, N. (2003). Dynamic latent factor models for intensity processes. CORE DP 2003/103.

Bauwens, L. and Rombouts, J. (2003). Bayesian clustering of many GARCH models. CORE DP 2003/87.

Bauwens, L. and Veredas, D. (2004). The stochastic conditional duration model: a latent factor model for the analysis of financial durations. Forthcoming in *Journal of Econometrics*.

Bhat, C. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B*, 35:677–693.

Bollerslev, T., Engle, R., and Nelson, D. (1994). ARCH models. In Engle, R. and McFadden, D., editors, *Handbook of Econometrics*, chapter 4, pages 2959–3038. North Holland Press, Amsterdam.

Brownstone, D. and Train, K. (1999). Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics*, 89:109–129.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90:1313–1321.

Chib, S. and Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, 85:347–361.

Chib, S. and Hamilton, B. (2000). Bayesian analysis of cross-section and clustered data treatment models. *Journal of Econometrics*, 97:25–50.

Chib, S., Nardari, F., and Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, 108:291–316.

Chintagunta, P. and Honore, B. (1996). Investigating the effects of marketing variables and unobserved heterogeneity in a multinomial probit model. *International Journal of Research in Marketing*, 13:1–15.

Cowles, M. and Carlin, B. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91:883–904.

Cribari-Neto, F. (1997). Econometric programming environnments: Gauss, Ox and S-plus. *Journal of Applied Econometrics*, 12:77–89.

Danielson, J. (1994). Stochastic volatility in asset prices: estimation with simulated maximum likelihood. *Journal of Econometrics*, 61:375–400.

de Jong, P. and Shephard, N. (1995). The simulation smoother for time series models. *Biometrika*, 82:339–350.

Deb, P. and Trivedi, P. (1997). Demand for medical care by the elderly: A finite mixture approach. *Journal of Applied Econometrics*, 12:313–336.

Diebolt, J. and Robert, C. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society, Series B*, 56:363–375.

Durbin, J. and Koopman, S. (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika*, 84:669–684.

Durbin, J. and Koopman, S. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of the Royal Statistical Society B*, 62:3–56.

Franses, P. and Paap, R. (2001). *Quantitative Models in Marketing Research.* Cambridge University Press, Cambridge.

Frühwirth-Schnatter, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*, 96:194–209.

Frühwirth-Schnatter, S. and Kaufmann, S. (2002). Bayesian clustering of many short time series. Working Paper, Vienna University of Economics and Business Administration.

Galli, F. (2003). Econometria dei Dati Finanzari ad Alta Frequenza. Tesi di Dottorato in Economia Politica, Bologna.

Geweke, J., Keane, M., and Runkle, D. (1997). Statistical inference in the multinomial multiperiod probit model. *Journal of Econometrics*, 80:125–165.

Ghysels, E., Harvey, A., and Renault, E. (1996). Stochastic volatility. In Maddala, G. and Rao, C., editors, *Handbook of Statistics*, pages 119–191. Elsevier Science, Amsterdam.

Gourieroux, C. and Monfort, A. (1997). *Simulation-based Econometric Methods.* Oxford University Press, Oxford.

Hajivassiliou, V. and Mc Fadden, D. (1998). The method of simulated scores for the estimation of LDV models. *Econometrica*, 66:863–896.

Hajivassiliou, V. and Ruud, P. (1994). *Classical estimation methods for LDV models using simulation.* In: Handbook of Econometrics, Volume 4, Chapter 40, North Holland, Amsterdam.

Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57:357–384.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58:83–90.

Horowitz, J. (1998). *Semiparametric Methods in Econometrics.* Springer Verlag, Berlin.

Jacquier, E., Polson, N., and Rossi, P. (1994). Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business and Economic Statistics*, 12:371–417.

Jacquier, E., Polson, N., and Rossi, P. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. Forthcoming in *Journal of Econometrics.*

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies*, 65:361–393.

Liesenfeld, R. and Richard, J.-F. (2003). Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of Empirical Finance*, 10:505–531.

Maddala, G. (1983). *Limited-dependent and Qualitative Variables in Econometrics.* Cambridge University Press, Cambridge.

Mariano, R., Schuermann, T., and Weeks, M. (2000). *Simulation-based Inference in Econometrics.* Cambridge University Press, Cambridge.

McCulloch, R., Polson, N., and Rossi, P. (2000). A Bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics*, 99:173–193.

McFadden, D. and Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15:447–470.

Paap, R. and Franses, P. (2000). A dynamic multinomial probit model for brand choice with different long-run and short-run effects of marketing-mix variables. *Journal of Applied Econometrics*, 15:717–744.

Pagan, A. and Ullah, A. (1999). *Nonparametric Econometrics*. Cambridge University Press, Cambridge.

Richard, J.-F. (1998). Efficient high-dimensional Monte Carlo importance sampling. Manuscript, University of Pittsburgh.

Richardson, S. and Green, P. (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, Series B*, 59:731–792.

Russell, J. (1999). Econometric modeling of multivariate irregularly-spaced high-frequency data. Manuscript, University of Chicago.

Sandman, G. and Koopman, S. (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics*, 67:271–301.

Shephard, N. (1996). *Statistical aspects of ARCH and stochastic volatility,* , chapter 1, pages 1–67. Time Series Models: In Econometrics, Finance and Other Fields. Chapman & Hall, in D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Nielsen (eds.), London.

Strickland, C., Forbes, C., and Martin, G. (2003). Bayesian analysis of the stochastic conditional duration model. Manuscript, Monash University.

Tanner, M. and Wong, W. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82:528–540.

Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge.

Index